PadhAI: Activation Functions & Initialization Methods

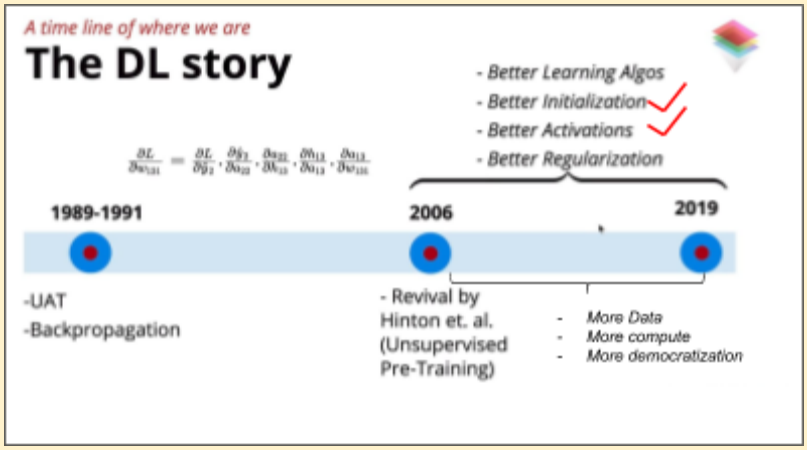
One Fourth Labs

Activation Functions

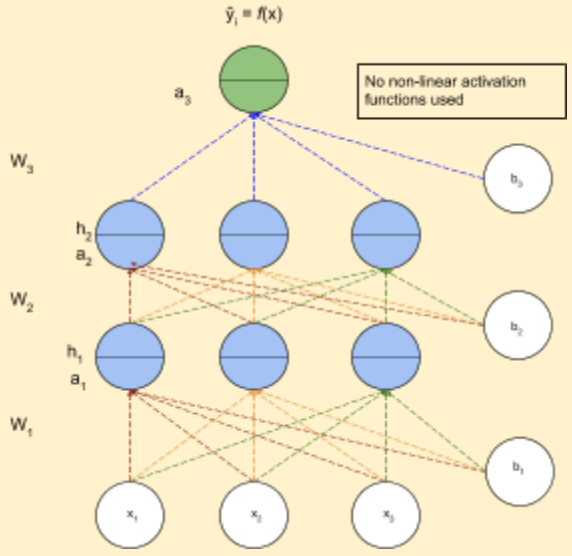
Setting the context

The progress of DL over the past decade

1. In this section, we will be looking at how better activation functions and better weight initialization has sped up the growth of DL over the last decade



2. Why are activation functions important?
   a. Consider a network where there are no non-linear activation functions like sigmoid etc.



   b. Here $\hat{y}_i = W_3(W_2(W_1(x_i))) = Wx_i$
   c. It can only represent linear relations between x and y
   d. Universal Approximation Theorem does not hold good.
   e. The **representation power** of a deep NN is due to its **non-linear activation functions**
3. Some popular non-linear activation functions are
   a. Logistic - can be called a sigmoid function
   b. tanh - can be called a sigmoid function
   c. ReLU
   d. Leaky ReLU